

EXHIBIT 3

(12) **United States Patent**
Circenis

(10) **Patent No.:** **US 6,816,809 B2**
(45) **Date of Patent:** **Nov. 9, 2004**

(54) **HARDWARE BASED UTILIZATION
METERING**

5,654,905 A 8/1997 Mulholland
6,049,798 A * 4/2000 Bishop et al. 707/10

(75) Inventor: **Edgar Circenis**, Loveland, CO (US)

FOREIGN PATENT DOCUMENTS

EP 0320329 6/1989

(73) Assignee: **Hewlett-Packard Development
Company, L.P.**, Houston, TX (US)

OTHER PUBLICATIONS

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 64 days.

Search Report issued on Jan. 14, 2004 in counterpart foreign
application in GB under application No. 0316794.7.

* cited by examiner

Primary Examiner—Bryan Bui
Assistant Examiner—Hien Vo

(21) Appl. No.: **10/200,175**

(22) Filed: **Jul. 23, 2002**

(65) **Prior Publication Data**

US 2004/0019456 A1 Jan. 29, 2004

(51) **Int. Cl.⁷** **G04F 1/00**

(52) **U.S. Cl.** **702/178; 702/186; 710/107;
710/260; 709/200**

(58) **Field of Search** 702/178, 186,
702/176; 710/260, 107, 241, 100; 709/104,
102, 107, 1, 103, 106, 226, 208, 200

(56) **References Cited**

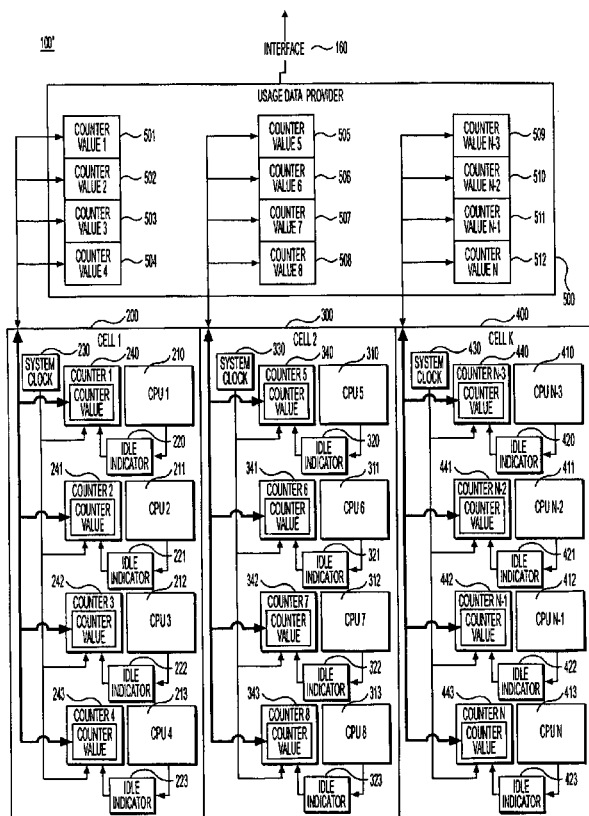
U.S. PATENT DOCUMENTS

4,503,495 A * 3/1985 Boudreau 710/241
4,511,960 A * 4/1985 Boudreau 711/219

(57) **ABSTRACT**

A hardware based utilization metering device, and a corresponding method are used in a computer system having one or more central processor units (CPUs) to provide a measure of CPU utilization. The device includes a state indicator coupled to a CPU. The state indicator receives an indication when the CPU is in a first state, which may be a busy state. A counter coupled to the state indicator and coupled to a system clock, receives a measure of system time from the system clock and receives data related to the indication when the CPU is in the first state, and generates a counter value indicative of time the CPU is in the first state. A data usage provider coupled to the counter maintains a non-volatile value of the counter value.

20 Claims, 6 Drawing Sheets



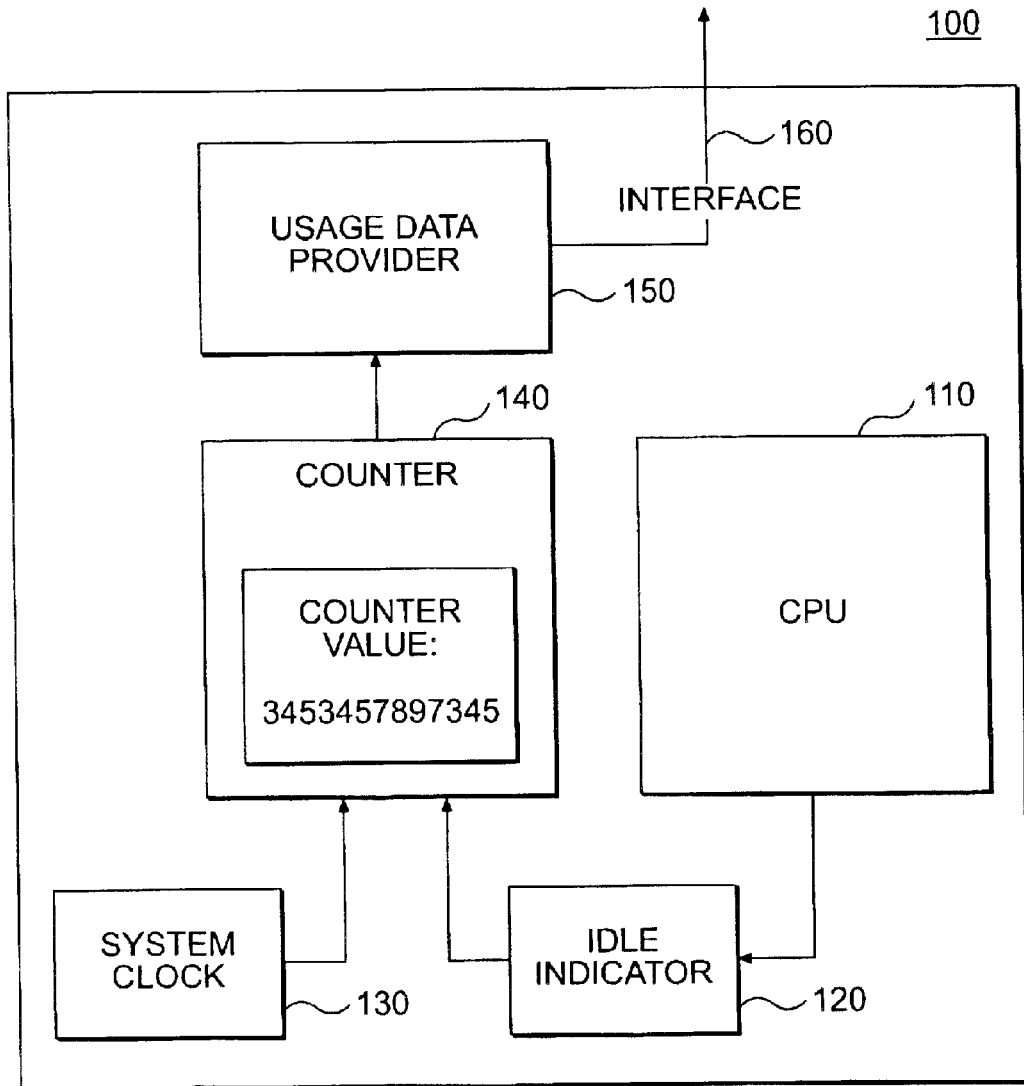


FIG. 1A

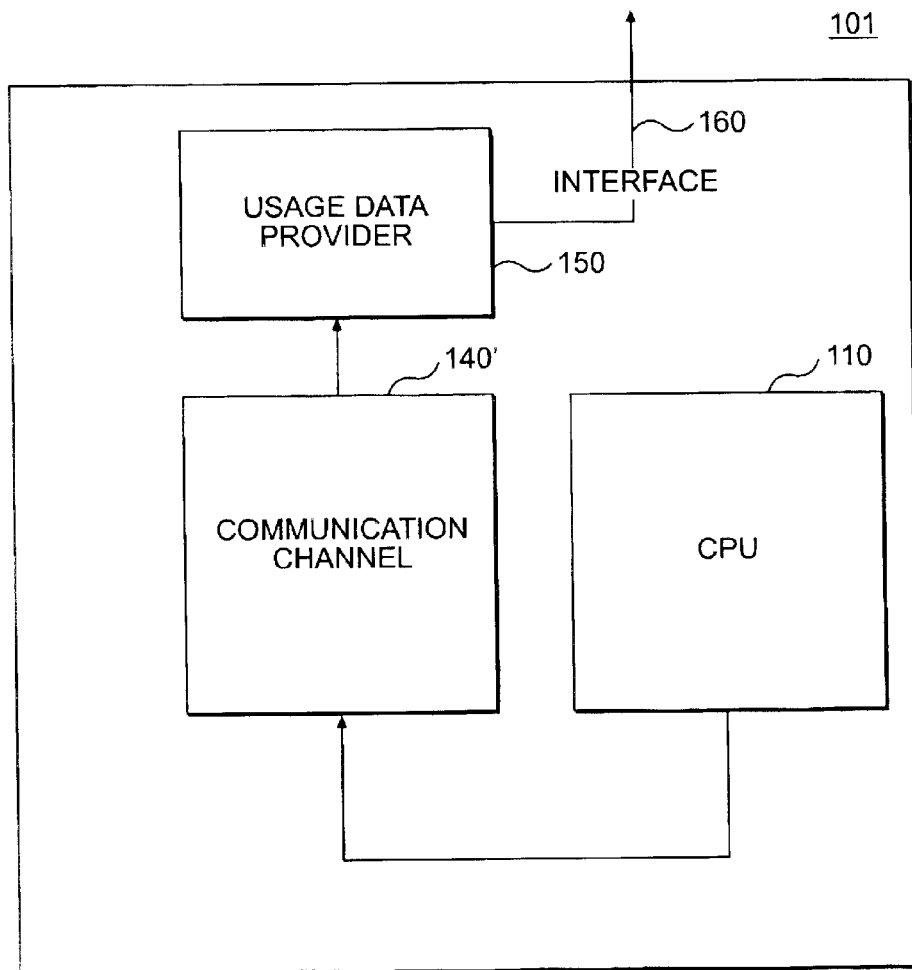


FIG. 1B

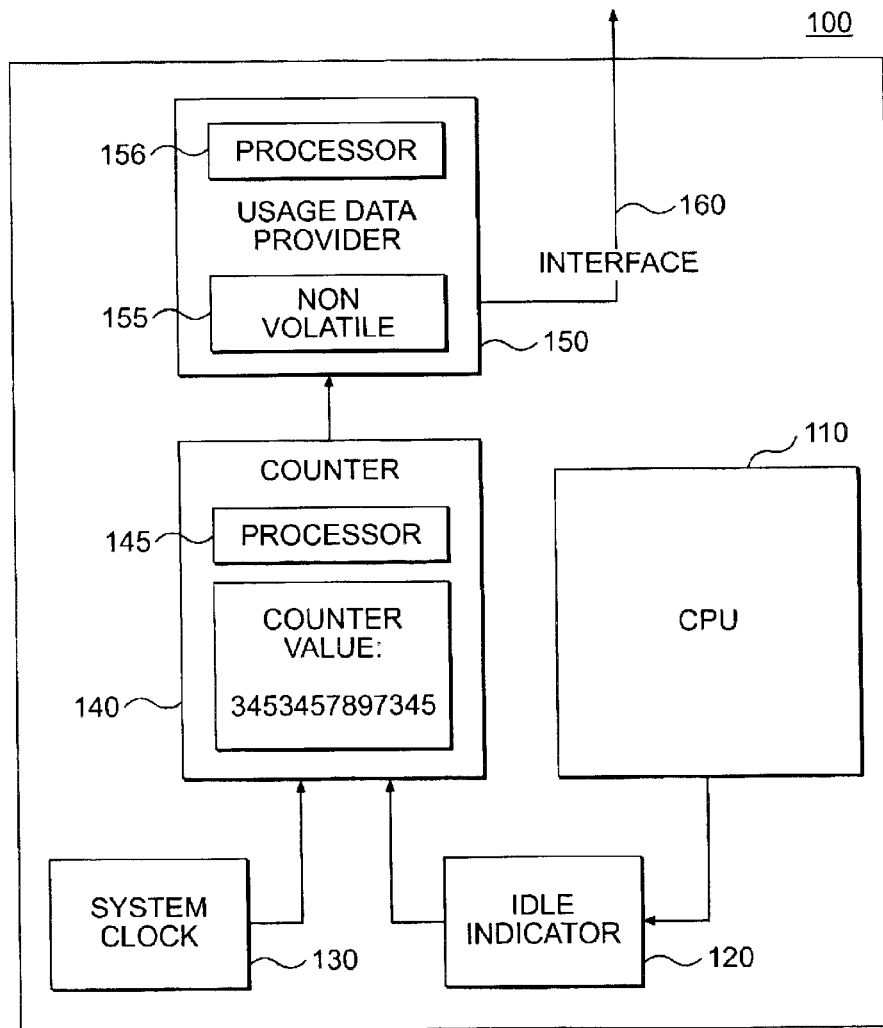
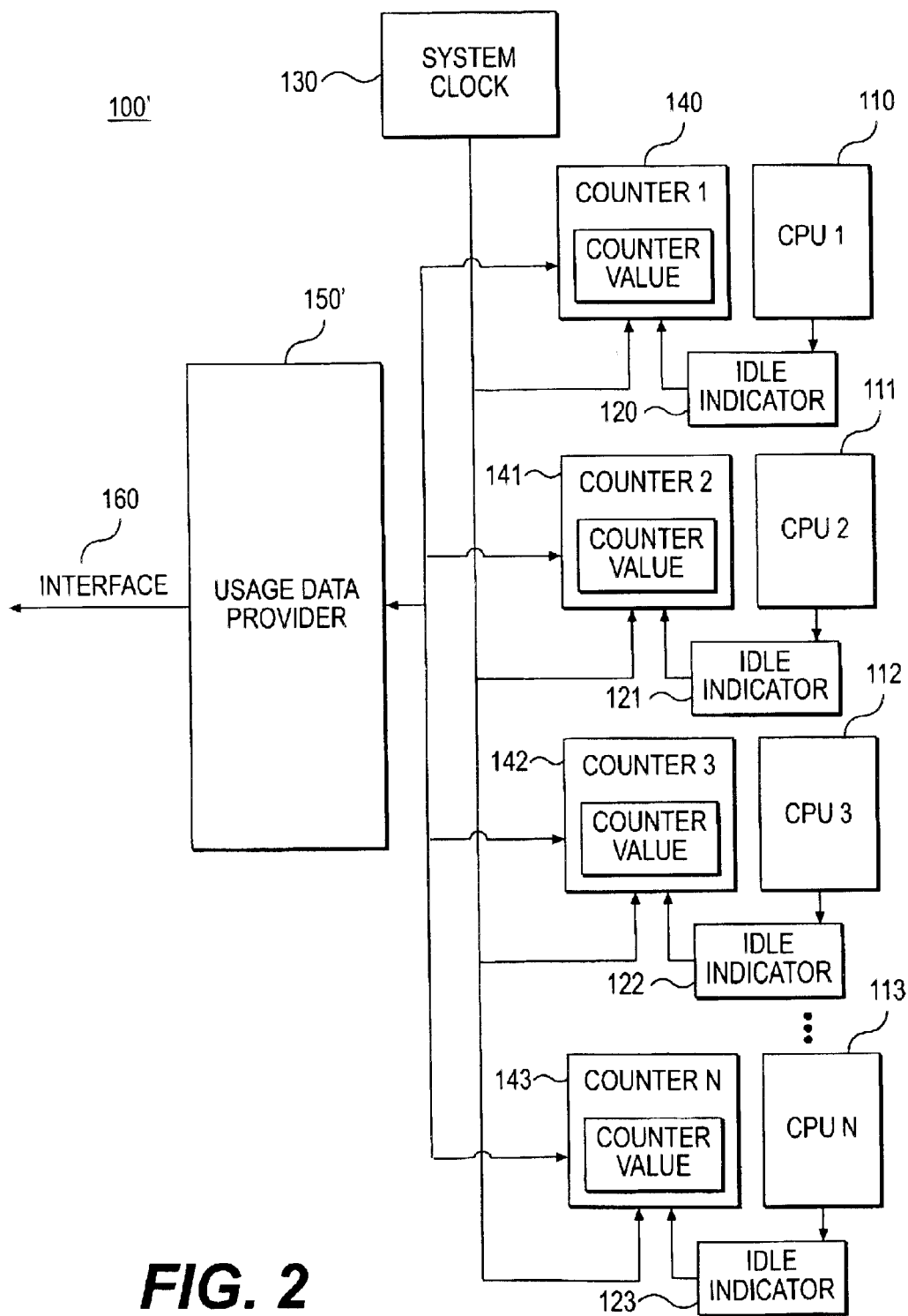
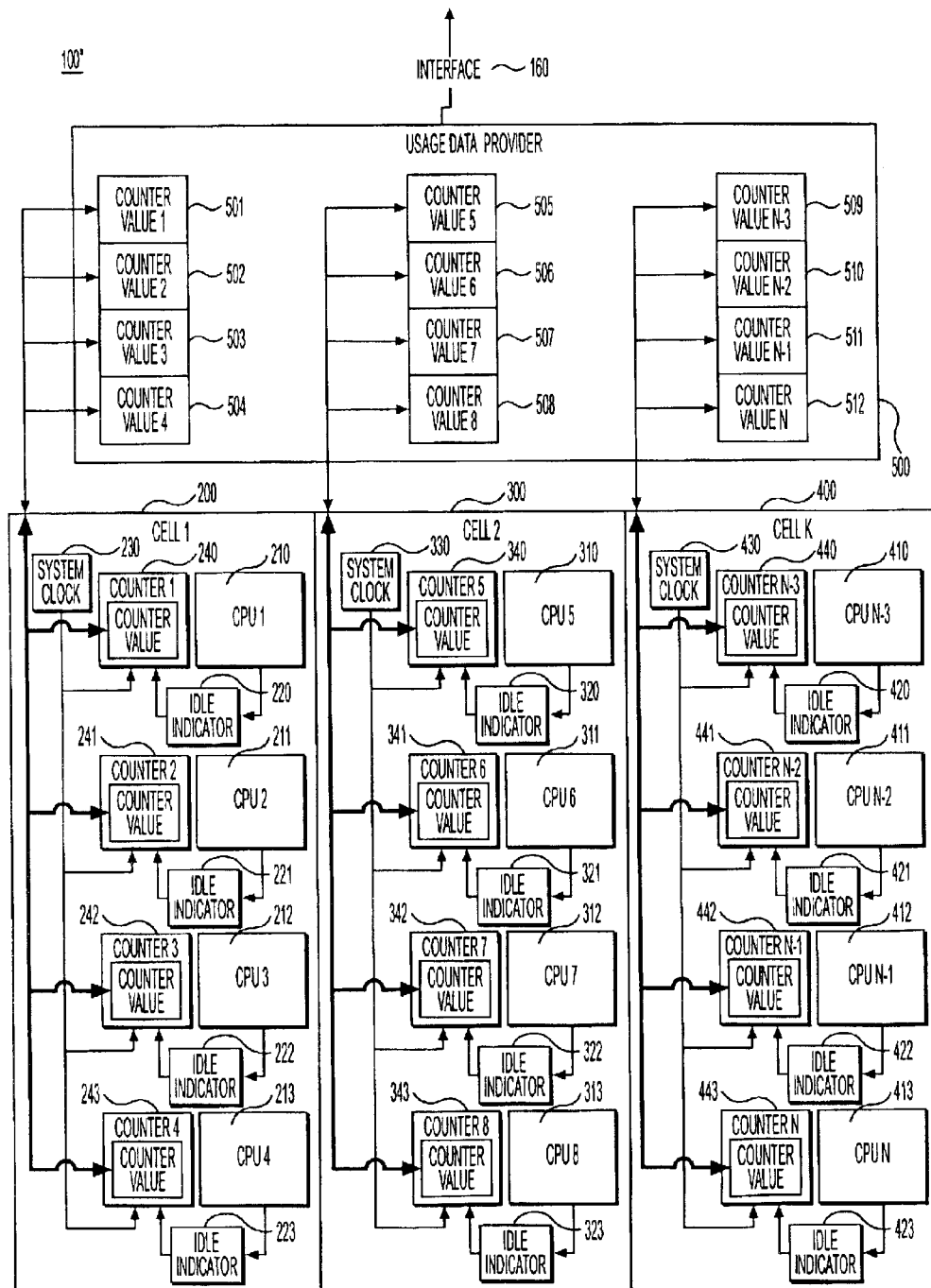


FIG. 1C



**FIG. 3**

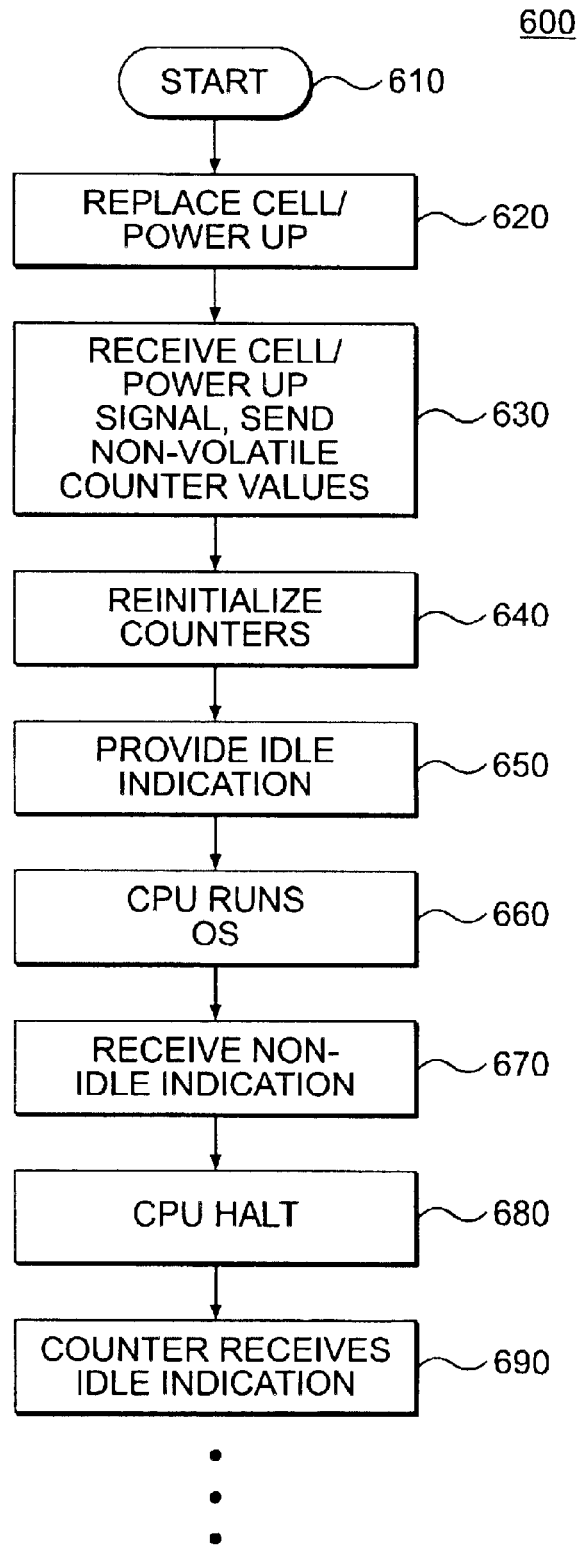


FIG. 4

US 6,816,809 B2

1

**HARDWARE BASED UTILIZATION
METERING****TECHNICAL FIELD**

The technical field is pay-per-use systems and methods that use central processor metering to determine processor utilization for billing and other purposes.

BACKGROUND

Many computer system users are turning to a pay-per-use concept in which the user is billed based on some measure of processor utilization within the computer system. A measure of processor utilization may be based on a metering scheme that determines when a processor is in use. Such metering of processor utilization is currently accomplished by software running within the computer system's operating system. In a computer system having hardware that may be partitioned, gathering processor utilization data from a hardware system requires communications between the metering application and all operating systems running within the hardware. The need for communication with different operating systems poses significant challenges because operating systems by their design are separated from other operating systems and do not have visibility to utilization data from other operating systems. Furthermore, the way partitioned systems are deployed, network connectivity may not be possible between different operating systems on the same hardware, further complicating the ability to accurately measure processor utilization. Current solutions involve communicating with each operating system independently, and then aggregating the information at a later time. This solution may be troublesome when network access is restricted, agent software is not installed, and operating systems are temporarily out of service, for example. This solution also requires the development of operating system-specific agents because each partition is capable of running different instances of the operating systems. In particular, since operating systems within a partitionable computer system do not, by default, communicate with each other, data providers (software agents) are needed to run on each operating system to collect and transmit, or otherwise provide utilization data from each partition. The thus-collected utilization data are aggregated to provide an overall value for processor utilization.

SUMMARY

What is disclosed is a hardware based utilization metering device for use in a computer system having one or more central processor units (CPUs), the device comprising a state indicator coupled to a CPU, wherein the state indicator receives an indication when the CPU is in a first state; a counter coupled to the state indicator and coupled to a system clock, wherein the counter receives a measure of system time from the system clock and receives data related to the indication when the CPU is in the first state, and generates a counter value indicative of time the CPU is in the first state; and a data usage provider coupled to the counter, wherein the data usage provider is capable of providing the counter value.

Also disclosed is a hardware based method for measuring processor utilization in a computer system comprising a plurality of processors, the method comprising determining when any of the plurality of processors is busy (i.e., not idle); providing a busy indication to a counter associated with the busy processor; receiving at the counter a measure

2

of computer system time; incrementing a counter value in the counter based on the received busy indication and an amount of computer system time that the processor is determined to be busy; and maintaining the counter value.

Finally, what is disclosed is an apparatus that provides hardware based utilization metering of CPUs in a computer system, comprising a plurality of CPUs. Associated with the CPUs is means for measuring computer system time. In addition, for each of the plurality of CPUs, the apparatus includes means for determining when the processor is busy; means for providing a busy indication when the processor is busy; means for receiving the busy indication and a measure of computer system time; means for combining the busy indication and the measure of computer system time to generate a counter value indicative of processor utilization; and means for maintaining the counter value.

DESCRIPTION OF THE DRAWINGS

The detailed description will refer to the following figures, in which like numbers refer to like elements, and in which:

FIG. 1A is a basic block diagram of a system that meters CPU utilization data in a computer system running multiple instances of operating systems;

FIG. 1B is a basic block diagram of an alternate system that meters CPU utilization data in a computer system running multiple instances of operating systems;

FIG. 1C is a more detailed block diagram of specific components of the system of FIG. 1A;

FIG. 2 is a further block diagram of a system that meters CPU utilization data in a computer system having multiple CPUs running multiple instances of operating systems;

FIG. 3 is yet a further block diagram of a system that meters CPU utilization data in a computer system having multiple CPUs running multiple instances of operating systems; and

FIG. 4 is a block diagram of a method for metering CPU utilization in a computer system having multiple CPUs running multiple instances of operating systems.

DETAILED DESCRIPTION

Pay-per-use systems allow computer users to acquire a given computing capacity that may be tailored to the user's specific need by, among other things, charging the user only for actual utilization of the system central processor units (CPUs). A measure of CPU utilization may be based on a metering scheme that determines when a processor is in use or busy. Such metering of CPU utilization is currently accomplished by software running within the computer system's operating system. This approach of CPU metering using software that runs within the operating system is acceptable for CPUs running single instances (i.e., type or version) of operating systems. For example, such a scheme would be acceptable for a computer system running only UNIX or only Windows®. However, for computer systems that may be hardware and software partitionable, gathering CPU utilization data from a hardware system requires communications between the metering application and all operating systems running within the hardware. The need for communication with different operating systems poses significant challenges because operating systems by their design are separated from other operating systems and do not have visibility to utilization data from other operating systems. Furthermore, the way partitioned systems are deployed, network connectivity may not be possible

US 6,816,809 B2

3

between different operating systems on the same hardware, further complicating the ability to accurately measure CPU utilization. Current solutions involve communicating with each operating system independently, and then aggregating the information at a later time. This solution may be troublesome when network access is restricted, agent software is not installed, and operating systems are temporarily out of service, for example. This solution also requires the development of operating system-specific agents because each partition is capable of running different instances of the operating systems. In particular, since operating systems within a partitionable computer system do not, by default, communicate with each other, data providers (software agents) are needed to run on each operating system to collect and transmit, or otherwise provide utilization data from each partition. The thus-collected utilization data are aggregated to provide an overall value for CPU utilization.

The disclosed apparatus and method for collecting CPU utilization data overcomes these problems. The apparatus and method assume that a CPU may be in a first state or in a second state. The CPU utilization may be based on a measure of time that the CPU spends in one state or the other. In an embodiment, the first state may be a busy state, generally meaning the CPU is running processes that perform useful work for the user of the computer system. Alternatively, the first state may be an idle state wherein the CPU is not performing useful work. The CPU utilization may be based on detecting or determining when the CPU is in the first state. In the embodiment in which the first state is the busy state, the apparatus may include means for determining if the CPU is busy. The means for determining if the CPU is busy may include hardware means. Alternatively, the means for determining if the CPU is busy may include software means. The apparatus also includes means for measuring and providing system time, and means for combining the system time and CPU busy indication to provide a CPU utilization value. In an embodiment, the CPU utilization may be a counter value. For example, a counter may increment by 1 for each system clock cycle that the CPU is not in an idle state. The hardware means for determining if the CPU is busy may include hardware modifications to the computer system. The alternate software means for determining if the CPU is idle may include modifications to the CPU's operating system(s). The apparatus may further include means for storing a non-volatile version of the CPU utilization value, means for initializing the CPU utilization value upon power up of the CPU, and means for reporting the CPU utilization value to a system or network external to the computer system.

FIG. 1A is a basic block diagram of a computer system **100** that implements hardware based utilization metering. The system **100** includes a CPU **110** on which multiple instances of operating systems may run. That is, the CPU **110** may support different types of operating systems (e.g., UNIX, Windows®), and different versions of a specific type of operating system. Coupled to the CPU **110** is an idle indicator **120**. The idle indicator **120** is capable of providing either an "idle" indication or a "not-idle"/"busy" indication. In an embodiment, the idle indicator **120** provides an output that indicates if the CPU **110** is "idle." The operation and structure of the idle indicator **120** will be described later. The output of the idle indicator **120** is provided to a counter **140**. The counter **140** also receives an output from a system clock **130**. The system clock **130** may measure ticks or cycles, or any other measure of system time, and then provide this data, or a modified form of the system time data, to the counter **140**.

4

Using the inputs from the system clock **130** and the idle indicator **120**, the counter **140** measures CPU cycles for the CPU **110**, where the CPU **110** is not in an idle state, but instead is performing a service for the user of the system **100**. While the CPU **110** is powered on, the counter **140** may thus maintain a counter value as shown in FIG. 1A, with the counter value (e.g., CPU cycles) updated based on the system time data and the idle indicator output.

The counter value is provided from the counter **140** to a usage data provider **150**. The counter value may be provided to the usage data provider **150** on a periodic basis or when the counter **140** is polled by the usage data provider **150**. The usage data provider **150** tracks the counter value and maintains a non-volatile master copy of the counter value. When the CPU **110** is powered on (or a hardware component containing the CPU **110** is powered on), the saved non-volatile counter value is provided from the usage data provider **150** to the counter **140** to initialize the counter value in the counter **140**. Because the usage data provider **150** maintains a non-volatile copy of the counter value, even if the CPU **110**, or other hardware component is removed, in addition to a loss of power situation, an up-to-date, or nearly up-to-date value of the counter value is always available.

The usage data provider **150** maintains a connection, or network interface **160** to a system or network (not shown) that is external to the computer system **100**. For example, the interface **160** may be a local area network (LAN) interface to a LAN. The LAN may include a management server that receives and processes information from the various computer systems coupled to the LAN, including the counter values that indicate CPU utilization. The usage data provider **150** can provide the current value of the counter value to the network by way of the network interface **160**. The counter value may be provided periodically or when polled by the network.

As noted above, the idle indicator **120** provides an indication that the CPU **110** is not idle. The idle indicator may be implemented as a hardware modification to the computer system **100**. For example, some CPUs include a pin on the CPU chip that provides a halt (idle) indication. Some operating systems halt the CPU when the CPU is not processing commands (i.e., the CPU is idle), and a halt (idle) indication (i.e., a high or low, or 0 or 1, value) may be asserted at the pin. In an embodiment, the idle indicator **120** may be coupled to the pin to read the halt (idle) indication. Other operating systems do not halt the CPU when the CPU is idle. Instead, the operating system may place the CPU into an idle loop, where the CPU remains until the operating system requires CPU processing. In an embodiment, a change to the operating system may be made such that upon entry into the idle loop, an externally visible register value is set to indicate the CPU is idle. When the CPU exits the idle loop, the register would be cleared. In addition to the just-described two embodiments of the idle indicator **120**, other structures and methods may be used to indicate the CPU **110** is idle.

FIG. 1B shows an alternative **101** to the embodiment to the hardware based utilization metering components shown in FIG. 1A. In FIG. 1B, the CPU **110** may run an operating system having a communications channel **140'** to the usage data provider **150**. The communication channel **140'** may be used to communicate CPU utilization metrics for collection by the usage data provider **150**. In this embodiment, hardware modification to meter the CPU **110** directly would not be required.

FIG. 1C shows selected components of the hardware based utilization metering device in more detail. In FIG. 1C,

US 6,816,809 B2

5

the counter **140** is shown including an optional processor **145** that may be used to process the idle indication and system time data to produce a CPU utilization metric in addition to the counter value. For example, the processor **145** may compute cumulative real or system time the CPU **110** is not idle. The usage data provider **150** is shown including a non-volatile storage **155** that stores the received counter value from the counter **140**. The non-volatile storage **155** may be any non-volatile storage device that is capable of receiving and storing the counter values. The usage data provider **150** may also include an optional processor **156** that may be used to convert the received counter value into a CPU metric such as total CPU non-idle hours, in real time, for example. The processor **156** may also handle requests for the utilization data over the interface **160** or for periodically transmitting utilization data to an external system on the LAN connected to the interface **160**.

FIG. 2 is a block diagram of a computer system **100'** that meters CPU utilization data where the computer system **100'** includes multiple CPUs running multiple instances of operating systems. In FIG. 2, the computer system **100'** includes CPU **1** through CPU **N** (denoted as **110–113**). This arrangement is meant to indicate that the computer system **100'** may include **N** CPUs, where **N** is an integer. Associated with each of the CPUs **110–113** is a corresponding idle indicator (**120–123**, respectively) and a corresponding counter (**140–143**, respectively). Each of the counters **140–143** receives an input from its respective idle indicator and from the system clock **130**. The system clock **130** shown in FIG. 2 performs the same functions, and has the same structure, as the system clock **130** shown in FIG. 1A. Each of the counters provides an output (counter value for its respective CPU) to a usage data provider **150'**. The usage data provider **150'** is similar to the usage data provider **150** shown in FIG. 1A, except that the usage data provider **150'** maintains a non-volatile measure of the counter value for each of the CPUs **110–113** in the computer system **100'**. Finally, the computer system **100'** includes the interface **160** for communicating CPU utilization data (counter values) to a system or network (not shown) external to the computer system **100'**.

FIG. 3 is a block diagram of a computer system **100''** having multiple CPUs arranged in partitions or cells. One or more of the CPUs may run multiple instances of operating systems, or, certain CPUs may run specific instances of operating systems while other CPUs run other instances of operating systems. As shown in FIG. 3, the computer system **100''** includes hardware cells **1** through **K**, denoted as cells **200**, **300**, and **400**, respectively. This arrangement is meant to indicate that the computer system **100''** may include **K** cells, where **K** is an integer, including **1** (one), indicating the computer system **100''** comprises one cell. Each of the cells is a set of components that can be removed from the computer system **100''** as a group. As can be seen in FIG. 3, the arrangement of components within each cell is identical in terms of type and number of components. However, the cells need not include the same type or same number of components. In addition to division of components among the hardware cells **1** through **K**, the components may be partitioned, or logically sorted. The partitions may comprise any number of CPUs or any number of cells. For example, cell **1** (**200**) may comprise four separate partitions, one for each of the CPUs installed in cell **1**. Alternatively, cell **1** (**200**) and cell **2** (**300**) may comprise a first partition and the remaining cell(s) may comprise additional partition(s).

Taking the cell **200** as an example, there are included four CPUs **210–213**. However, the cell **200** may include more or

6

fewer than four CPUs. Associated with each of the CPUs **210–213** is a corresponding idle indicator **220–223**. Each of the idle indicators receives an indication when its associated CPU is idle, and provides an output to a corresponding counter **240–243** that provides a current counter value (CPU utilization metric). Each of the counters **240–243** receives an input from a system clock **230**. The system clock **230** functions in the same manner as the system clock **130** shown in FIG. 1A.

Each of the counters **240–243** provides its associated counter value to a usage data provider **500**. The usage data provider **500** maintains a copy of the counter value (shown in FIG. 3 as **501–512**, respectively) for each of the CPUs in the cells **200**, **300** and **400**. The usage data provider **500** includes the interface **160** to a system or network (not shown) external to the computer system **100''**.

The arrangement of the cells **200**, **300**, and **400** allows one or more of the cells to be removed from the computer system **100''** while maintaining the computer system in operation. To prevent loss of the counter values for the CPUs in a removed cell, the counter values are stored in the usage data provider **500**. In an embodiment, the stored counter values may be maintained in non-volatile storage. In operation, the counter values that are closest to the CPUs (i.e., the counter values in the counters) are incremented, and the counter values in the usage data provider **500** are periodically updated. The counter values may be updated based on a specific reporting interval, or by a polling action initiated by the usage data provider **500**.

Should one of the cells **200**, **300** or **400** be replaced with a new cell, or with the original cell, but with one or more new CPUs, the counter values may be lost when the cell is powered off. Thus, whenever a cell is powered on, the usage data provider **500** will reinitialize the corresponding counters in the cell. For example, if the cell **200** is removed, and the CPU **210** is replaced with a new CPU **210'**, then when the original cell **200** is reinstalled and powered on, the usage data provider **500** will provide the stored counter value **501** to the counter **240**. The usage data provider **500** will also initialize the counters **241–243** by providing the stored counter values **502–504**, respectively. Because the CPU **210** was replaced by the new CPU **210'**, the provided counter value **501** will reflect the CPU utilization for the CPU **210**. Subsequent non-idle time of the CPU **210'** will be counted by incrementing the counter **240**.

In the above example, the replacement CPU **210'** is similar in all respects to the original CPU **210**, so that incrementing the counter **240** with the new CPU **210** installed should proceed the same as with the original CPU **210**. However, the new CPU **210** may also differ in some respects from the original CPU **210**. In an embodiment, the apparatus and method for metering CPU utilization may include the necessary means for accounting for differences between CPUs. For example, a processor in the counter **240** (similar to the processor **145** shown in FIG. 1C) or a processor in the usage data provider **500** (similar to the processor **156** shown in FIG. 1C) may adjust the cumulative counter value for CPU **210** should CPU **210** be replaced with a different type or model CPU, or should the system clock **230** change clock rate, or should an operating system running on the CPU **210** change such that determination of the counter value previously determined at the counter **240** would change. For example, should the system clock **230** be replaced with a system clock having a higher clock rate than that of the original system clock **230**, the processor in the counter **240** could simply multiply the determined counter value by the ratio of the original clock rate to the new clock

US 6,816,809 B2

7

rate to produce a consistent rate of change in the counter value from that determined using the original clock rate and that determined using the new clock rate.

FIG. 4 is a flow chart showing a hardware based CPU utilization operation 600 using the system 100" of FIG. 3. The operation 600 starts in block 610. In block 620, the cell 200 is replaced and the hardware components on the cell 200 are powered up. In block 630, the usage data provider 500 receives an indication of the power up of the cell 200 components, and the usage data provider 500 sends the stored counter values 501-504 to the counters 240-243, respectively. In block 640, the received counter values are used to reinitialize the counters 240-243. However, the CPUs 210-213 are idle (operating systems are not processing), and in block 650, the idle indicators 220-223 provide an idle indication to their respective counters 240-243, thereby preventing incrementing of the counter values. In block 660, the CPU 210 begins running an operating system, and the idle indicator 220 sends an indication to the counter 240 that the CPU 210 is not idle. In block 670, the counter 240 receives the non-idle indication, and the system time from the system clock 230, and begins incrementing the counter value for the CPU 210. In block 680, the operating system running on the CPU 210 stops processing, halts the CPU 210, and asserts a halt indication. In block 690, the counter 240 receives an idle indication and stops incrementing the counter value. The operation 600 may continue with incrementing counter values for other CPUs in the system 100" and may include routines to update the non-volatile counter values in the usage data provider 500.

What is claimed is:

1. A hardware based utilization metering device, comprising:

- an idle indicator coupled to a processor, wherein the idle indicator receives an indication when the processor is in a first state;
- a counter coupled to the idle indicator and coupled to a system clock, wherein the counter receives a measure of system time from the system clock and receives data related to the indication when the processor is in the first state, and generates a counter value indicative of time the processor is in the first state; and
- a data usage provider coupled to the counter, wherein the data usage provider is capable of providing the counter value.

2. The device of claim 1, wherein the first state is a busy state.

3. The device of claim 2, wherein the processor comprises an operating system that halts the processor when the processor is not processing, and wherein the idle indicator is a hardware device coupled to a pin of the processor, wherein the hardware device reads a signal asserted on the pin to receive the idle indication.

4. The device of claim 2, wherein the processor comprises an operating system that places the processor in an idle loop when the processor is idle, and wherein the idle indicator comprises an externally visible register that holds a value indicating the processor is in the idle state.

5. The device of claim 2, wherein the counter value increments based on the measure of system time when the processor is in the busy state.

6. The device of claim 1, wherein when the processor is powered on, the counter receives a current value of the counter value.

7. The device of claim 1, wherein the counter provides the counter value to the data usage provider on a periodic basis.

8. The device of claim 1, wherein the usage data provider polls the counter to obtain the counter value.

9. The device of claim 1, further comprising an interface coupling the data usage provider to an external network, the

8

interface used for providing processor utilization data to the external network.

10. The device of claim 1, further comprising a plurality of processors, each of the plurality of processors having an associated idle indicator and an associated counter, whereby the associated idle indicator and the associated counter combine to provide a counter value for their respective processor.

11. The device of claim 10, wherein the usage data provider maintains a non-volatile counter value for each of the plurality of processors.

12. The device of claim 10, wherein the plurality of processors are arranged in one or more cells.

13. A hardware based method for measuring processor utilization in a computer system comprising a plurality of processors, the method comprising:

- determining when any of the plurality of processors is busy;
- providing a busy indication to a counter associated with a busy processor;
- receiving at the counter a measure of computer system time;
- incrementing a counter value in the counter based on the provided busy indication and an amount of computer system time that the processor is determined to be busy; and

maintaining the counter value.

14. The method of claim 13, further comprising:

- reinitializing the counter value in the counter, using the maintained value of the counter value, when the processor is powered on.

15. The method of claim 13, wherein maintaining the counter value comprises:

- updating in a usage data provider a current value of the counter value from the counter; and
- storing the updated current counter value.

16. The method of claim 15, wherein updating the current counter value comprises periodically receiving the current counter value from the counter.

17. The method of claim 15, wherein updating the current counter value comprises polling the counter.

18. An apparatus that provides hardware based utilization metering of central processor units (CPUs) in a computer system, comprising:

- a plurality of CPUs, wherein the plurality of CPUs are arranged in one or more cells;
- for each cell, means for measuring computer system time; and

for each of the plurality of CPUs:

- means for determining when a CPU is busy;
- means for providing an indication when the CPU is busy;
- means for receiving the indication and measure of computer system time,
- means for combining the indication and the measure of computer system time to generate a counter value indicative of CPU utilization, and
- means for maintaining the counter value.

19. The apparatus of claim 18, further comprising means for reinitializing the counter value upon power of the CPU.

20. The apparatus of claim 18, further comprising means for reporting a CPU utilization value to a network external to the computer system.

* * * * *